ISSN 22773061

Sparsity based Single Object Tracking

Glincy Abraham, K.A.Narayanankutty, K.P.Soman
M.Tech, Final Year Student
Department of CEN
Amrita Vishwa Vidyapeetham, Coimbatore
glincy01@gmail.com

Professor
Department of ECE
Amrita Vishwa Vidyapeetham, Coimbatore

Professor and Head
Department of CEN
Amrita Vishwa Vidyapeetham, Coimbatore

ABSTRACT

Object tracking has importance in various video processing applications like video surveillance, perceptual user interface driver assistance, tracking etc. This paper deals with a new tracking technique that combines the dictionary based background subtraction along with sparsity based tracking. The speed and performance challenges faced during the sparsity based tracking alone are addressed, as it is based on a background subtraction preprocessing and local compressive tracking. It also overcomes the challenges faced by the traditional techniques due to illumination variation, pose and shape change of the object. Output of the proposed technique is compared with that of compressive tracking technique.

Indexing terms/Keywords

Background subtraction, compressive tracking

Council for Innovative Research
Peer Review Research Publishing System
Journal: INTERNATIONAL JOURNAL OF COMPUTERS & TECHNOLOGY
Vol 9, No 2
editor@cirworld.com
www.cirworld.com, member.cirworld.com

1. INTRODUCTION

Computer vision is a wide area that deals with various methods for acquiring, processing, examining high-dimensional data from the real world for making useful decisions about real world physical objects and scenes. With the development of surveillance cameras, the amount of data that is collected through these cameras and the area covered by them are skyrocketed. Availability of low cost video surveillance devices in turn increases use of video surveillance application in various fields. Visual tracking is the process of predicting the location of an object in an image sequence over time. Video tracking is a very useful task for video surveillance that has been used in security sensitive areas for observing and monitoring behavior activities or other change in information.

Tracking creates a projection trajectory of the moving object. Tracking algorithms can be classified into two, mainly as generative and discriminative [1]. Both classifications are performed based on the target appearance model. The generative model describes the appearance of the object with the help of a template [2] or by using simple subspace. In order to adapt to the appearance of the target, template of target is updated dynamically [3]. Similarly in case of subspace model, subspace which represents the target is learned continuously during the tracking process. Discriminative algorithms pose the tracking problem as a binary classification task in order to find the decision boundary for separating the target object from the background. Features that discriminates between the background and the object are taken for object tracking. In discriminative tracking method, various features are selected and are learned online for tracking [4].

The commonly used steps in the process of object tracking includes, identifying the objects of interest in the video sequence and to cluster pixels of these objects. There are many surveys available about background subtraction [5] [6] and object tracking for video surveillance [7] [8]. The survey presented here gives an overview about traditional methods.

1.1 Background Subtraction

Background subtraction (BGS) creates the foreground mask which holds the location of foreground objects. Based on the background and foreground modeling, background subtraction method can be mainly classified into two categories, pixel-level methods and frame level methods. In pixel-level methods, the distribution of the pixel in the frame is considered both locally and independently. One example of pixel level method is the frame differencing [9]. Frame differencing is a fast method but is not able to preserve interior pixel information of a uniformly colored moving object. A running Gaussian average model was proposed by C. Wren in [10]. Here the background is modeled independently at each pixel location. A Gaussian probability is ideally positioned over the last *n* pixel values. A running average is computed to avoid fitting of probability density function in each frame. Another method used for object detection is Gaussian mixture model(GMM) [11]. A GMM is a parametric probability density function represented as a weighted sum of Gaussian component densities. Nir Friedman and Stuart Russell used mixture of Gaussian method for traffic surveillance system [12]. In a paper, Stauffer and Grimson [13] extend the idea of Friedman using multiple Gaussians. Non- parametric model [14] can give better result to backgrounds having clutters and moving scenes such as swaying trees, bushes. In nonparametric model, model structure is specified from the data itself. A Kalman filter method models distribution of each pixel in the background, and for a new pixel. The probability is calculated for finding out whether it belongs to background. In [15], algorithm is developed to segment out foreground object from the dynamically varying background. Dynamic texture background model is created and Kalman filter algorithm is used for iteratively estimating the appearance texture model as well as the foreground object.

Pixel-level BGS methods suffers from certain deficiencies as this method ignores the spatial relationship between the pixels in the background. But this geometric information is necessary for proper background modeling and foreground segmentation.

Different from pixel level model, frame level model considers the frame as a whole image and the variation in the background image is considered as foreground objects. In [16] foreground object detection is done by eigen background subtraction. The main goal of the system is to cluster the pixels into blobs based on motion, since we have a stationary background with moving objects. In [17], Markov random field is introduced to impose group sparsity to foreground pixels. Here, the background subtraction was done by compressive sensing method. Background subtracted images shows a spatial sparsity in the foreground object regions. An alternative approach for background subtraction based on compressive sensing is dynamic group sparsity [18]. Dynamic group sparsity is a greedy sparse recovery algorithm. This algorithm prunes the foreground object by using an iterative process based on both sparsity and group clustering instead of only sparsity as in various other methods.

1.2 Object Tracking

Video surveillance activities can be manual, semi-automatic or automatic. In manual surveillance system, the analysis was done by human. Such systems are widely used. Both human and system have equal importance in semi-automatic system. Simple motion detection [19] systems are example for such systems. Automatic systems perform full video processing without human interaction.

Karan Gupta et al [20], presented a work for employing automatic object tracking based on the concepts of both template matching and frame differencing. In this work, frame differencing is used in a frame-by-frame basis for finding moving object with preferably high speed. In [21], object tracking based on particle filter is explained. Particle filter represents the probability distribution over various states as a set of weighted samples or particles. Each particle represents the possible location of the object to be tracked. During particle filtering, large number of particles are used to predict the posterior probability.

Speeded Up Robust Features (SURF) [22] is a feature descriptor which generates a set of feature points that are invariant to geometric transformation like rotation, scaling etc. Instead of selecting random points, in [23] SURF generated points are taken as initial samples to particle filter algorithm. In [24], particle filter algorithm is used along with the features obtained from Scale-invariant feature transform (SIFT) descriptors. But SURF is faster than SIFT, so it gives a better result with better speed. Mean shift is an iterative non- parametric feature space analysis technique used for clustering analysis in computer vision. In this algorithm color histogram is used to represent target. Continuously adaptive mean shift (CamShift) algorithm [25] based on an adaptive mean shift is a low complex reliable tracking algorithm.

In this paper, Dictionary based background subtraction [26] is combined with the sparsity based object tracking algorithm [27]. This combination of two algorithms provides a better result for tracking. Dictionary based background subtraction and compressive tracking technique is explained in Section 2 and Section 3 respectively. In Section 4 proposed technique is explained. Section 5 handles the result and its discussion. Conclusion and future work is explained in Section 6.

2. DICTIONARY BASED BACKGROUND SUBTRACTION

Background subtraction (BGS) is a critical step for various computer vision applications like video surveillance, traffic monitoring, human action recognition etc. Main steps involved in background subtraction are background modeling and foreground detection. Background modeling is the most challenging step in BGS. Major problems faced during background modeling are the natural changes in the background throughout the video and the presence of foreground objects in the training sample.

Dictionary based BGS [26] process requires two stages of operation, training stage and foreground detection stage. Frame work of dictionary based BGS is shown in figure 1 [26]. Here the background modeling is done with the help of dictionary learning

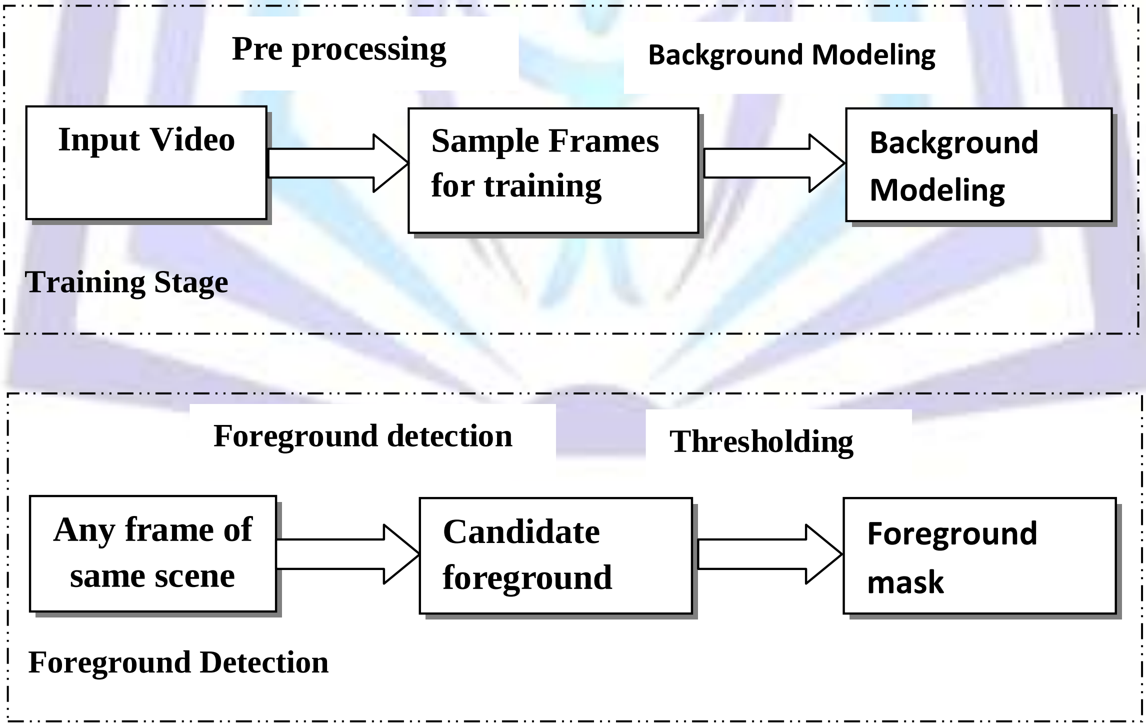


Fig 1: Framework of dictionary based background subtraction

2.1 Background Modeling

A given frame can be represented as a combination of background and foreground as shown in the equation.

$$x = x_b + x_f \tag{1}$$

This methodology is based on two sparsity assumptions. Firstly, background of a given frame has a sparse representation over the learned dictionary D.

$$x_b = D\alpha \tag{2}$$

α is the co efficient vector. After background subtraction, foreground objects are sparse over background.

$$x_f = x - D\alpha \tag{3}$$

The background subtraction problem can be written as

$$\alpha = \arg\min_{\alpha} \|x - D\alpha\|_1 + \lambda \|\alpha\|_1 \tag{4}$$

In dictionary based background subtraction technique, training samples are collected randomly from the original video itself and the basis for the dictionary is formed from these samples. This dictionary is learned directly, and the optimal learned dictionary is obtained by solving the equation:

$$D = \arg\min_{D,A} \|X - DA\|_1 + \lambda.\|A\|_1 \tag{5}$$

X is a matrix containing eigenvectors corresponding to highest eigen values of training sample as column. A is the coefficient matrix. The coefficient matrix as well as the dictionary needs to be optimized. This optimization problem is done in two steps

(1) Optimize the coefficient matrix A by keeping D as constant

$$A = \arg\min_{A} \|X - DA\|_1 + \lambda \|A\|_1 \tag{6}$$

(2) Update D by keeping A as constant

$$D = \arg\min_{D} \|X - DA\|_1 \tag{7}$$

A and D are updated iteratively by keeping the other as constant. These two steps are repeated to get an optimized solution for A and D

2.2 Foreground Segmentation

Pixels deviating from the background are considered as foreground candidates. Second sparsity consideration is used here for getting foreground candidates. Probability of a pixel in x_f belonging to foreground depends both on its pixel value and also its neighborhood.

$$Score(i) = x_f^{\,2}(i) + \sum_{j \in neighbor(i)} x_f^{\,2}(j) \tag{8}$$

A threshold is applied on the score for creating the foreground mask.

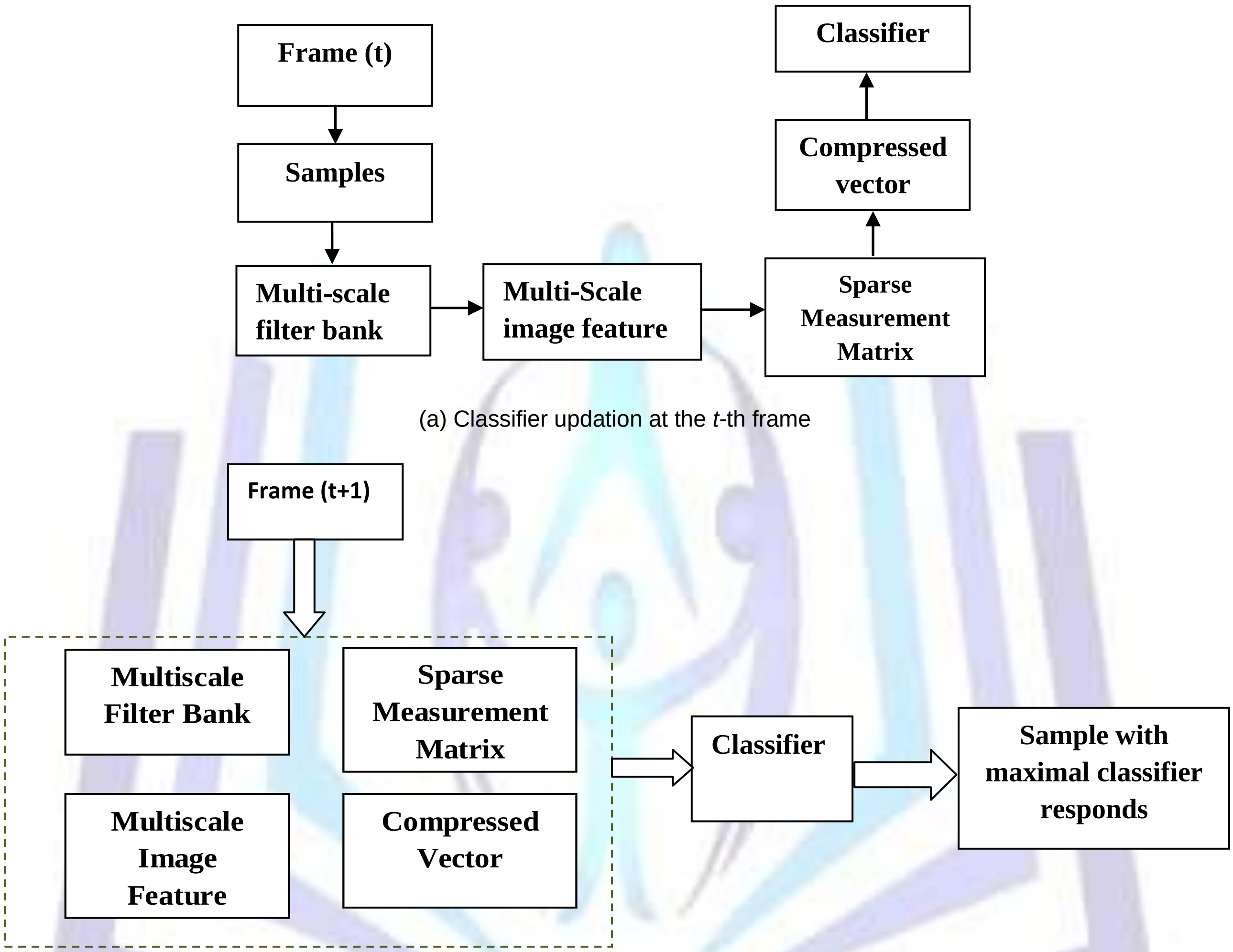
3. COMPRESSIVE TRACKING

Compressive tracking algorithm is based on the features extracted from compressive domain. This tracking method is generative, since the appearance model is obtained from feature extracted in compressive domain, as well as discriminative because the target detection is obtained via features applied on Naive Bayes classifier. In Figure 2 [29], shows the main components of the compressive tracking method.

Tracking window is selected from the first frame. From each frame some positive and negative samples are taken. Positive samples are taken from the locations near to the target and negative samples are taken further away from the location. These samples are used for updating the classifier. In order to find the target location in subsequent frames, maximum classification score is estimated from the samples taken from the nearby location of the current target. Sample with maximum classification score is considered as the target.

4. PROPOSED METHOD

Proposed method combines the dictionary based background subtraction method along with compressive tracking. In case of compressive tracking, classification is done for samples taken from nearby locations of current target. In the case of complex backgrounds, chance for misclassification is high. In order to avoid this situation we are introducing background subtraction technique before tracking process as a preprocessing step. Video data obtained from a surveillance camera is used as input to this algorithm. Once the preprocessing is over, certain number of frames are selected randomly to train the system. From the selected training samples, few samples are taken and eigen decomposed. Eigenvector corresponding to highest eigen values are stacked together to form initial dictionary. This dictionary is updated by solving the optimization problem mentioned in section 2. complex background is removed and foreground candidates alone is available. In this case we are not going for post processing step in background subtraction. Frame containing foreground candidate alone is given as input to compressive tracking algorithm. Compressive tracking algorithm works in similar way as explained in section 3. Output of proposed technique gives a better result than that of compressive tracking technique.



(b). Tracking at the $(t + 1)$th frame

Fig 2 : Main Components of Compressive tracking

5 RESULTS AND DISCUSSION

The proposed method is implemented in matlab and experimented on two different videos, human and train. For the evaluation of proposed method, output of proposed method is compared with output of compressive tracking method. Output of background subtraction method is shown in figure 3. Eigen vectors corresponding to three highest eigen value for each randomly selected frame is used for initial dictionary creation. Frame number is specified in the upper left corner of the image. Output of compressive tracking on three different frame of the video, human, is shown in a, b, c of figure 4. Output of proposed method is shown in d, e, f of figure 4. Figure g to l of figure 4 shows the output of both the algorithms in second video. Tracking performance evaluation is done with the help of visual observation.

ISSN 22773061

(a) (b)

Fig 3: Background subtraction (a) original frame (b) foreground objects obtained after background subtraction.

(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Fig 4: a, b, c, g, h, i shows output of compressive tracking algorithm. d, e, f, j, k, l are output obtained from proposed method.

6. CONCLUSION

In this paper we are combining a sparsity based technique for background subtraction and object tracking. Selected background subtraction method, dictionary based BGS, gives improved object detection than previous methods. This method gives superior results even for changing background. In this paper, we are proposing an effective tracking technique that will give improved time and tracking performance than previous methods. This method gives improved result for complex video sequences and performs well in terms of robustness and accuracy.

ACKNOWLEDGMENTS

I am highly grateful to Assistant Professor Dr. Geetha P for providing reviews on the project. I would also like to express my gratitude towards my parents for their kind encouragement which helped me to accomplish this paper work.

REFERENCES

1. Qian Yu, Thang Ba Dinh, and Gerard Medioni ; “Online Tracking and Reacquisition Using Co-trained Generative and Discriminative Trackers”. ECCV 2008, Springer-Verlag Berlin Heidelberg (2008)
2. Iain Matthews, Takahiro Ishikawa, and Simon Baker; “The Template Update Problem”. IEEE Transactions on Pattern Analysis and Machine Intelligence , Vol 26: (810–815), 2004.
3. David A. Ross , Jongwoo Lim, Ruei-Sung Lin, Ming-Hsuan Yang; “Incremental Learning for Robust Visual Tracking”. IJCV, 77:125–141(2008).
4. Helmut Grabner, Michael Grabner and Horst Bischof; “Real-Time Tracking via Online Boosting”. BMVC, 47–56 (2006)
5. Zane Mayo and Jules R Tapamo; ”Background Subtraction Survey for Highway Surveillance”. proceedings/prasa09-14(2009)
6. U.Chandrasekhar ,Tapan kumar Das;” A Survey of Techniques for Background Subtraction and Traffic Analysis on Surveillance Video”. ISSN 2250-0987 , UNIASCIT, Vol 1 , 107-113(2011)
7. C. Lakshmi Devasena , R. Revathí and M. Hemalatha; “Video Surveillance Systems – A Survey”. IJCSI, Vol 8, Issue 4, July 2011
8. Kinjal A Joshi, Darshak G. Thakore;” A Survey on Moving Object Detection and Tracking in Video Surveillance System”. IJSCE, ISSN: 2231-2307, Vol 2, Issue 3, July 2012
9. Widyawan, Muhammad Ihsan Zul; “Adaptive Motion Detection Algorithm using Frame Differencesand Dynamic Template Matching Method”. URAI 2012)page 26-28, 2012.
10. C. Wren, A. Azarhayejani, T. Darrell, and A.P. Pentland; “Pfinder: Real-Time Tracking of The Human Body,” IEEE Trans. on Patfern Anal. and Machine Infell, vol. 19, page 780-785, (1997).
11. Douglas A. Reynolds; “Gaussian Mixture Models”. Encyclopedia of Biometrics 2009: 659-663, (2009)
12. N. Friedman and S. Russell; “Image Segmentation in Video Sequences: A Probabilistic Approach”. UAI 97, August 1997.
13. Stauffer C, Grimson W. E. L. “Adaptive Background Mixture Models for Real-Time Tracking”. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.
14. A. M. Elgammal, D. Harwood, and L. S. Davis; “Nonparametric Model for Background Subtraction”. Proceedings of ECCV ’00, page 751–767, 2000.
15. C. Stauffer and W. E. L. Grimson; “Adaptive Background Mixture Models for Real-Time Tracking”. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’99), page 246–252, June 1999.
16. N. M. Oliver, B. Rosario, and A. P. Pentland; “A Bayesian Computer Vision System for Modeling Human Interactions”. IEEE Transactions on Pattern Analysis andMachine Intelligence, vol. 22, page. 831–843, 2000.
17. V. Cevher,M. F. Duarte, C.Hegde, and R. G. Baraniuk; “Sparse Signal Recovery Usingmarkov Randomfields”.NIPS ’08, page 257–264, 2008.
18. J. Huang, X. Huang, and D. Metaxas; “Learning with Dynamic Group Sparsity”. ICCV ’09, page. 64–71, Oct 2009.
19. S. Zhu, and A. Yuille; “Region Competition: Unifying Snakes, Region Growing, and Bayes/Mdl for Multiband Image Segmentation”. IEEE Trans. Patt. Analy. Mach. Intell. 884–900, (1996).
20. Karan Gupta, Anjali V. Kulkarni; “Implementation of an Automated Single Camera Object Tracking System Using Frame Differencing and Dynamic Template Matching”. Book title = "SCSS (1)'07", PAGES (245-250),(2007).

ISSN 22773061

21. Lyudmila Mihaylova, Paul Brasnett, Nishan Canagarajah and David Bull, "Object Tracking by Particle Filtering Techniques in Video Sequences". Advances and Challenges in Multisensor Data and Information.pages (260-268)(2007)

22. H. Bay, T. Tuytelaars and L. Van Gool, "SURF: Speeded Up Robust Features", in: 9th European Conference on Computer Vision, (2006).

23. H. Kandil and A. Atwan. "A Comparative Study between SIFT- Particle and SURF-Particle Video Tracking Algorithms". International Journal of Signal Processing, Image Processing and Pattern Recognition Vol. 5, Sept, (2012)

24. B. Feng, B. Zeng and J. Qiu, "Tracking Object by Combining Particle Filters and SIFT Features", in: IEEE Fifth International Conference on Image and Graphics, page. 527-532 (2009).

25. Chunrong Zhang, Yuansong Qiao, Enda Fallon, Chiangqiao Xu "An Improved CamShift Algorithm for Target Tracking in Video Surveillance". 9th. IT & T Conference, School of Computing, Oct 23 (2009)

26. Cong Zhao, XiaogangWang, andWai-Kuen Cham. "Background Subtraction via Robust Dictionary Learning", EURASIP Journal on Image and Video Processing,(2011).

27. Kaihua Zhang, Lei Zhang, and Ming-Hsuan Yang "Real-Time Compressive Tracking". ECCV 2012, LNCS 7574, pp. 866–879, Springer-Verlag Berlin Heidelberg (2012).